Geophysical Research Abstracts
Vol. 15, EGU2013-**PREVIEW**, 2013
EGU General Assembly 2013

Per pixel uncertainty modelling and its spatial representation on land cover maps obtained by hybrid classification.

Xavier Pons (1), Eva Sevillano (1), Gerard Moré (2), Pere Serra (1), Dan Cornford (3), and Miquel Ninyerola (4)

(1) Department of Geography, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain (xavier.pons@uab.cat), (2) CREAF, Cerdanyola del Vallès, Spain (g.more@creaf.uab.cat), (3) Computer Science, Aston University, Birmingham, United Kingdom (d.cornford@aston.ac.uk), (4) Department of Botanic, Cerdanyola del Vallès, Spain (miquel.ninyerola@uab.cat)

The usage of remote sensing imagery combined with statistical classifiers to obtain categorical cartography is now common practice. As in many other areas of geographic information quality assessment, knowing the accuracy of these maps is crucial, and the spatialization of quality information is becoming ever more important for a large range of applications. Whereas some classifiers (e.g., maximum likelihood, linear discriminant analysis, naive Bayes, etc) permit the estimation and spatial representation of the uncertainty through a pixel level probabilistic estimator (and, from that, to compute a global accuracy estimator for the whole map), for other methods such a direct estimator does not exist. Regardless of the classification method applied, ground truth data is almost always available (to train the classifier and/or to compute the global accuracy and, usually, a confusion matrix). Our research is devoted to the development of a protocol to spatialize the error on a general framework based on the classifier parameters, and some ground truth reference data. In the methodological experiment presented here we provide an insight into uncertainty modelling for a hybrid classifier that combines unsupervised and supervised stages (implemented in the MiraMon GIS).

In this work we describe what we believe is the first attempt to characterise pixel level uncertainty in a two stage classification process. We describe the model setup, show the preliminary results and identify future work that will be undertaken.

The study area is a Landsat full frame located at the North-eastern region of the Iberian Peninsula. The six non-thermal bands + NDVI of a multi-temporal set of six geometrically and radiometrically corrected Landsat-5 images (between 2005 and 2007) were submitted to a hybrid classification process, together with some ancillary data (climate, slopes, etc). Training areas were extracted from the Land Cover Map of Catalonia (MCSC), a 0.5 m resolution map created by exhaustive photointerpretation of orthophotos from 2005-2007; a sampling of the 50% of the total MCSC surface was used to train the classifier. For validation and uncertainty estimation purposes, the two remaining 25% subsamples of the MCSC were used. The MCSC original legend was reduced from 233 categories to 21 categories.

A logistic regression was applied in order to estimate the probability of being successfully classified for each pixel. The dependent variable was a binary layer of correctly/incorrectly classified labels obtained from the test areas. The independent variables were extracted from different classifier parameters or computed during the supervised stage of the classifier. Examples of these variables are fidelity, representativity, entropy, uncertainty, promiscuity, etc.

Preliminary results show that a simple model containing the classifier parameter fidelity as the only significant independent variable performed best. The area under the receiver operating characteristic (AUROC) curve of 0.692 is very similar to the general accuracy estimated with the confusion matrix (0.685). The validation of this model (the estimation of a success or an error) shows a coherent accuracy of 0.693.

We believe this is the first work where the spatial uncertainty of a hybrid classification is accurately validated. Further steps will include refining the classification results, attempting to improve the uncertainty model and testing a per-polygon approach. Moreover, this protocol will be applied to other classifiers. In future work a standard classification scenario (ground truth not covering the whole area) will be applied in order to perform a sensitivity analysis of the protocol, and the results will be compared with a classical Bayesian classification approach.